Evolution of Gene Expression in the Cell-division Cycle of Woodland Populations of Budding Yeast

Daniel F. Simola
Laboratory of Junhyong Kim
29 Oct 2007

Penn University of Pennsylvania
Natural populations of yeast

- \textit{S. cerevisiae}: most important model organism for genetics and functional genomics?

- Laboratory strains domesticated, but likely derived from natural, \textbf{WILD} strains

- Little known about evolutionary history (genetic or phenotypic) of these natural yeast strains...
### Genetic diversity of *S. cerevisiae*

<table>
<thead>
<tr>
<th>Source</th>
<th>Within $\pi \ (e^{-3})$</th>
<th>Between $\pi \ (e^{-3})$, all ORFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saké wine</td>
<td>1.0</td>
<td>3.7</td>
</tr>
<tr>
<td>Grape wine</td>
<td>1.4</td>
<td>4.4</td>
</tr>
<tr>
<td>Clinical</td>
<td>4.2</td>
<td>5.6</td>
</tr>
<tr>
<td>Nature</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>Fermentation</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4.2</strong></td>
<td><strong>93.8</strong></td>
</tr>
</tbody>
</table>

Fay & Benavides, PLoS Genet., 2005

*S. paradoxus*
**ORF alignments of 40 cerevisiae strains**

- Gene ID by cat’ing Blast HSPs, and realigning with pairwise global N-W
- Toss ORFs with improper CDS format
- \( \pi = \text{Proportion of SNPs per base, averaged over sequence pairs} \)

<table>
<thead>
<tr>
<th>( \pi e^{-3} # )</th>
<th>CDS</th>
<th>Intron</th>
<th>5' 1kb</th>
<th>3' 1kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>10.0 (6464)</td>
<td>50 (99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Periodic</td>
<td>12.5 (775)</td>
<td>101.3 (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF</td>
<td>13.1 (266)</td>
<td>53.5 (6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosome</td>
<td>11.9 (441)</td>
<td>47.6 (31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meiosis</td>
<td>13.3 (105)</td>
<td>35.3 (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sporulation</td>
<td>12.5 (75)</td>
<td>29.3 (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo-hyphal</td>
<td>14.5 (60)</td>
<td>NA (0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Excess of noncoding diversity

<table>
<thead>
<tr>
<th>vs. Lab</th>
<th>ORF</th>
<th>Intronic</th>
<th>Flanking 500</th>
<th>UTRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vineyard RM11-1a</td>
<td>3.4</td>
<td>7.4</td>
<td>4.0</td>
<td>7.9</td>
</tr>
<tr>
<td>Clinical YJM789</td>
<td>4.1</td>
<td>17.7</td>
<td>9.2</td>
<td>8.5</td>
</tr>
<tr>
<td>Natural YPS606</td>
<td>5.2</td>
<td>9.3</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

---

**Pie Chart**

- **NC**: 22%
- **TNC**: 10%
- **P**: 68%

*Genome Biology*
Genetic variability among woodland isolates

- Introns from 9 unlinked loci sequenced (4.9 kb), reveals diversity/base of $1.89 \pm 0.62 \times 10^{-3}$
- Phylogeny supports 3 haplotypes with minimal within group and low between group diversity... little phenotypic diversity?
- Microsatellite sequencing supports this phylogeny

<table>
<thead>
<tr>
<th>ID</th>
<th>HO-Kan Strain ID</th>
<th>Origin</th>
<th>&quot;Neutral&quot; Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>YPS2067</td>
<td>Tyler Arboretum, PA</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>YPS2055</td>
<td>Tyler Arboretum, PA</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>YPS3137</td>
<td>Jenkins Woods, PA</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>YPS2060</td>
<td>Mettlers Woods, NJ</td>
<td>B</td>
</tr>
<tr>
<td>7</td>
<td>YPS2066</td>
<td>Mettlers Woods, NJ</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>YPS2079</td>
<td>Westtown School Woods, PA</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>YPS2073</td>
<td>Mettlers Woods, NJ</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>YPS3060</td>
<td>Jenkins Woods, PA</td>
<td>C</td>
</tr>
</tbody>
</table>
Sequencing at $G_1$ start

Jorgensen & Tyers, 2004 (modified)

Rupes, 2002
Genetic variability of CGP loci

- Total of 45 segregating sites over 5256 bases
- Averaging $9.92 \pm 2.4 \times 10^{-3}$ diversity/base
  
  $\text{cf. } 1.89 \pm 0.62 \times 10^{-3}$

| ID   | HO-Kan Strain ID | Origin       | “Neutral” Haplotype | cln3 3' | far1 5' | far1 3' | mbp1 5' | ifh1 5' | whi5 5' | whi3 5' | sfp1 3' | crf1 3' |
|------|------------------|--------------|----------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1    | YPS2067          | Tyler Arboretum, PA | A        | A       | A       | A       | C       | A       | A       | A       | A       | A       |
| 2    | YPS2055          | Tyler Arboretum, PA | A        | A       | A       | A       | A       | A       | A       | A       | A       | A       |
| 3    | YPS3137          | Jenkins Woods, PA  | A        | A       | A       | A       | A       | A       | A       | A       | A       | A       |
| 6    | YPS2060          | Mettlers Woods, NJ | B        | A       | B       | A       | B       | D       | A       | A       | A       | A       |
| 8    | YPS2079          | Westtown School Woods, NJ | B | B | B | B | B | B | B | B | B | B | ? |
| 4    | YPS2073          | Mettlers Woods, NJ | C        | C       | B       | A       | A       | A       | B       | B       | C       | C       |
| 5    | YPS3060          | Jenkins Woods, PA  | C        | C       | B       | A       | A       | ?       | C       | B       | C       | ?       |

Incomaptibilities with haplotype tree
Transcriptome diversity

• Genome alignments indicate an excess of noncoding variation among strains

• Targeted sequencing reveals elevated variation in promoter & UTRs relative to introns

How does this noncoding variation affect gene expression & cell cycle progression?
Previous studies

- Numerous studies characterize gene expression variability in lab yeast (2400+ slides published), but evolutionary studies are lacking

- Developmental time-series studies
  - Respiration (DeRisi 1997)
  - Sporulation (Chu 1998)
  - Cell-division cycle (Spellman 1998)
  - Environmental perturbation responses (Gasch 2000)

- Other studies investigated transcriptome evolution
  - *Drosophila melanogaster* subgroup (Rifkin 2003)
  - *C. elegans* isolates (Denver 2005)
  - *Drosophila* development (Rifkin 2005)

- No study to date has combined both approaches in a high-resolution manner (in any organism)
Experimental design

- HO: Kan transform diploid isolates, isolating MATα haploids
- Synchronize cultures with α-factor pheromone
- Upon release, sample cultures at 18 points (20 min) through 1.3 cell cycles, at 18°C in SD medium
- 2 dye-swapped technical replicates of each sample
- Unsynchronized lab strain common reference hybridized with each synchronized samples

36 slides per strain
Microarray data collection status

8 \textit{S. cerevisiae} isolates, spanning the 3 haplotype groups
• 300 microarrays

1 laboratory strain, S288c
• 36 microarrays

1 \textit{S. paradoxus} strain
• 36 microarrays

Complete collection for 9 strains, totaling 365 microarrays
Data set calibrations

- **Reference channel bias**
  correct for multiple biological samplings of reference channel cells

- **Morphological variation**
  calibrate measurements with respect to developmental time instead of clock time
Reference channel bias

- Biological batches of reference RNA may vary due to culturing conditions, freezer shelf time

- YPS2079
- YPS2060
- YPS3137
- · · ·
- YPS183

Nov 2005 - Dec 2006

- 65 genes > 1.39 fold
- 245 genes > 1.39 fold
Reference channel correction

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2007</th>
<th>2005</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y'  = \lg(M/R_{05}) + \lg(R_{05}/R_{07})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y'  = \lg(M/R_{06}) + \lg(R_{06}/R_{07})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Clock time ≠ developmental time

- Want to assess transcriptome variability at identical phenotypic states (developmental stages)
- But strain-specific variation in rate of developmental progression w.r.t. clock time necessitates calibration
- Estimate global scaling parameter as ratio of cycle lengths w.r.t. strain with longest period (lab strain) to avoid extrapolation
Average temporal expression
After filtering Ty elements
Ty elements

- LTR retrotransposons
- 5 classes, 90 genes
- Preferentially insert near Pol III transcribed genes (e.g. tRNA)
- Regulated primarily through post-transcriptional mechanisms
Average Ty element expression

![Graph showing the expression of Ty elements over time](image-url)

- **Lg(Exp/Ref) (Average over 92 genes)**
- **Mitotic index**
- **Time (min after release)**
Temporal expression variability

- Cell cycle progression controlled by dynamic TF regulation
- Regulation induces temporal ordering constraints, seen as checkpoints

Hypothesis: transcriptome comparison will reveal a signature of evolutionary history, which reflects these developmental constraints
Biological replicate variance

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variance \{ \log(‘05/’07) \ \log(‘06/’07) \}

BGSS
for each gene
Comparable to technical replicate error
H₀: Temporal noise

- Phase noise will produce variance $\propto (dx/dt)^2$

- Scale each gene's $F[t]$ by derivative$^2[t]$
Derivative estimation
Optimal lag times

Histogram of optimal lag times (per gene trajectory)
Lag deviations over strain
Temporal expression variability

Variability ($F_{t, Y_t^{1.2}}$)

- 114 Meiosis, 1/7.26
- 74 Sporulation, 1/2.75
- 59 Pseudohyphal, 1/0.96
- 485 Ribosome, 1/0.01
- 268 TF, 1/1.45
- 318 Periodic, 1/28.15

6066 Genes, 1/100000000

2 SE(Variance)


- 1/100
- 1/10000

Time (min after release)
Haplotype A variability

Temporal Expression Variability of Woodland Yeast Strains

- 114 Meiosis, 1/3.39
- 74 Sporulation, 1/10.55
- 59 Pseudohyphal, 1/0.74
- 485 Ribosome, 1/0.02
- 268 TF, 1/5.91
- 800 Periodic, 1/40.0

Variability ($F_{Y_{1}}^{2}$)

- 6128 Genes, 1/100000000
- 2 SE(Variance)


- 1/100
- 1/10000

Time (min after release)
Haplotype B variability

Temporal Expression Variability of Woodland Yeast Strains

- Variability (\(E_i Y_i^{-2}\))
  - 14 Meiosis, 1/20.18
  - 74 Sporulation, 1/4.7
  - 59 Pseudohyphal, 1/5.86
  - 485 Ribosome, 1/0.02
  - 268 TF, 1/1.44
  - 798 Periodic, 1/41.16

- 6159 Genes, 1/100000000
- 2 SE(Variance)

- Derivative Avg. Expr.
  - 1/100

  - 1/10000

Time (min after release)
Haplotype C variability

Temporal Expression Variability of Woodland Yeast Strains

Variability ($F_1Y_{t-1}$)

- 114 Meiosis, 1/9.65
- 74 Sporulation, 1/3.54
- 59 Pseudohyphal, 1/4.11
- 485 Ribosome, 1/0.06
- 268 TF, 1/1.74
- 800 Periodic, 1/26.98

6264 Genes, 1/100000000

2 SE(Variance)

Derivative Avg. Expr.

1/100

1/10000

Time (min after release)
Average($F$) = 0.004±0.014 (1/1e8 scale)
Overall transcriptome diversity

8 *S. cerevisiae* isolates, spanning the 3 haplotype groups

1 laboratory strain, *S288c*

1 *S. paradoxus* strain
Haplotype-specific functional diversity

- Do functional groups exhibit differential variation among haplotypes?

<table>
<thead>
<tr>
<th>ANOVA (n=9)</th>
<th>F-value</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>58.57</td>
<td>0.72</td>
</tr>
<tr>
<td>Periodic</td>
<td>60.14</td>
<td>1.97</td>
</tr>
<tr>
<td>TF</td>
<td>63.09</td>
<td>3.35</td>
</tr>
<tr>
<td>Ribosome</td>
<td>61.36</td>
<td>2.51</td>
</tr>
<tr>
<td>Meiosis</td>
<td>65.05</td>
<td>5.58</td>
</tr>
<tr>
<td>Sporulation</td>
<td>66.07</td>
<td>6.52</td>
</tr>
<tr>
<td>Pseudohyphal</td>
<td>65.57</td>
<td>7.39</td>
</tr>
</tbody>
</table>

$p < e^{-5}$ when $F > 22.59$
Most variable genes

- Which genes exhibit greatest variability among haplotypes?

<table>
<thead>
<tr>
<th>Gene</th>
<th>F-value</th>
<th>GO description</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC34</td>
<td>439.1</td>
<td>SCF complex; ubiquitin-protein ligase which regulates cell cycle progression by targeting key substrates for degradation</td>
<td>2</td>
</tr>
<tr>
<td>KAP114</td>
<td>437.5</td>
<td>Responsible for nuclear import of Spt15p, histones H2A and H2B, and Nap1p</td>
<td>3</td>
</tr>
<tr>
<td>ELP4</td>
<td>437.2</td>
<td>Elongator protein, component of the RNA polymerase II holoenzyme</td>
<td>4</td>
</tr>
<tr>
<td>SNZ2</td>
<td>434.1</td>
<td>Stationary phase-induced gene family; induced prior to diauxic shift, and also in the absence of thiamin in a Thi2p-dependent manner</td>
<td>5</td>
</tr>
<tr>
<td>SNF2</td>
<td>430.1</td>
<td>Catalytic subunit of the SWI/SNF chromatin remodeling complex involved in transcriptional regulation</td>
<td>6</td>
</tr>
<tr>
<td>ADH1</td>
<td>306.5</td>
<td>Alcohol dehydrogenase, required for the reduction of acetaldehyde to ethanol, the last step in the glycolytic pathway</td>
<td>37</td>
</tr>
</tbody>
</table>

FWER < .05 when F > 22.59
Variability of CGP loci

Do the sequenced CGP loci exhibit significant among haplotype transcriptome diversity?

<table>
<thead>
<tr>
<th>Gene</th>
<th>F-value</th>
<th>Rank</th>
<th>Woodland $\pi$</th>
<th>5’ <em>cerevisiae</em> $\pi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cln3</td>
<td>12.7</td>
<td>5469</td>
<td>5.82</td>
<td></td>
</tr>
<tr>
<td>Far1</td>
<td>18.9</td>
<td>4782</td>
<td>7.72</td>
<td></td>
</tr>
<tr>
<td>Mbp1</td>
<td>17.9</td>
<td>4893</td>
<td>1.52</td>
<td></td>
</tr>
<tr>
<td>Sfp1</td>
<td>32.5</td>
<td>3606</td>
<td>16.8</td>
<td></td>
</tr>
<tr>
<td>Whi3</td>
<td>14.0</td>
<td>5340</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td>Whi5</td>
<td>11.1</td>
<td>5602</td>
<td>14.8</td>
<td></td>
</tr>
<tr>
<td>Crf1</td>
<td>5.29</td>
<td>5943</td>
<td>22.0</td>
<td></td>
</tr>
<tr>
<td>Ifh1</td>
<td>20.4</td>
<td>4613</td>
<td>12.16</td>
<td></td>
</tr>
</tbody>
</table>

FWER < .05 when F > 22.59

4412 total sig. genes
Summary

• Sequencing of CGP loci reveals elevated genetic variability relative to intronic & microsatellite loci. *What is neutral?*

• Cross-strain transcriptome variability is a function of the cell cycle, peaking during growth phases

• Major life cycle gene groups exhibit among haplotype transcriptome variability, despite low genetic diversity
Acknowledgments

• Chantal Francis, transformations, sync. arrays
• Jeremy Weiss, imputation code
• Irmina Gawlas, sequencing, paradoxus arrays
Fungi: Ascomycota

Wapinski et al, Nature 2007